Global Canopy Atlas pipeline

Fabian Fischer et al.

2025-04-29

- 1 Accronyms
- 2 Package description
- 3 General workflow
- 4 Input data
- 5 Parameters
- 6 Usage
- 7 Output products
- 8 Recent developments
- 9 Current challenges
- 10 Future work

1 Accronyms

- **GCA**: Global Canopy Atlas
- **DTM**: Digital Terrain Model
- **DSM**: Digital Surface Model
- **CHM**: Canopy Height Model
- **ALS**: Aerial Lidar Scanning
- **CRS**: Coordinate Reference System
- **UTM**: Universal Transverse Mercator
- **TIN**: Triangle Irregular Network
- **MAAP**: Multi-Mission Algorithm and Analysis Platform
- **ESA**: European Space Agency

2 Package description

The GCA pipeline is a tool used for the generation of Digital Terrain

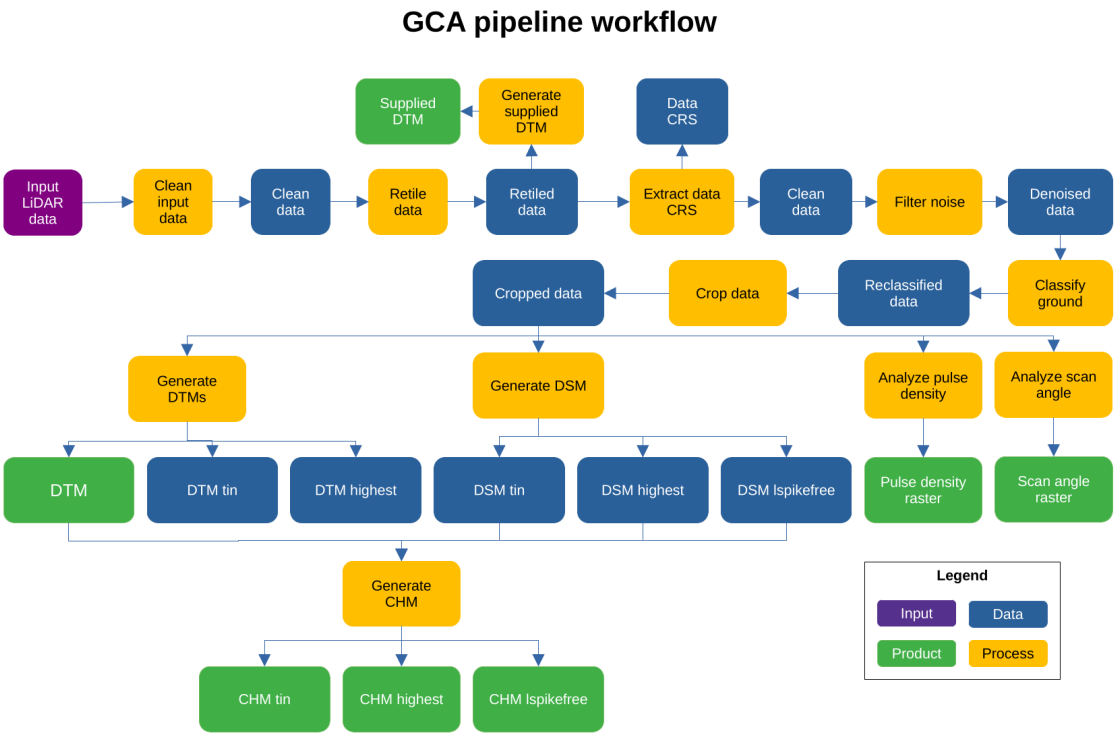
Model (DTM) and Canopy Height Model (CHM) from Aerial LiDAR Scanning (ALS) data.

The package source code can be found at: https://github.com/fischer-fjd/GCA/tree/GCA_open_source

3 General workflow

- **Diagram**

The following figure shows a diagram of the pipeline workflow.



In this figure, the blue items represent data being processed by each step, the orange items represent the processing steps of the pipeline and the green items represent the final output products of the pipeline.

- **Description**

TODO

4 Input data

The input data is provided in the form of a folder containing a set of pointcloud files, in format LAS (.las) or its compressed equivalent LAZ

(.laz). Normally each pointcloud file corresponds to a tile of the scanned landscape. All tiles must have the same Coordinate Reference System (CRS), which should be UTM compatible.

A metadata file should be provided with the data in the same folder. This file shall contain:

- The data CRS
- etc

5 Parameters

Name	Description	Default
name_job	Overall job name, used for processing stats	“gca”
type_file	Type of the files to be processed, needs to be exact (las, laz)	“las”
dir_dataset	Folder that contains data sets	“”
dir_processed	Folder where processed data sets should be saved	“”
path_lastools	Folder to most recent lastools installation	“”
tmpdir_processing	Folder where processing occurs (files will be overwritten)	“”
resolution	Resolution of raster products (in m)	1.0
n_cores	Number of cores for processing	1
size_tile	Retiling size	500
size_buffer	Tile buffer size	50

Name	Description	Default
force.utm	Force reprojection of coordinate reference system into UTM (and meter) coordinates	True
force.recompute	Force reprocessing. Only unprocessed data subsets will be reprocessed	False

6 Usage

- Place the input data in the desired input folder
- Complete and verify the input data metadata
- Fill the desired values for each parameter
- Run the main file: ALS_processing.R

7 Output products

- Supplied DTM
- DTM
- CHM
    - TIN
    - Highest
    - Lspikefree
- Pulse density raster
- Scan angle raster

8 Recent developments

- Open source – General functions
- Open source – Ground classification

9 Current challenges

- Finish the implementation of all open source functions

10 Future work

1. **Complete integration of the new ground classification algorithm in the pipeline**

   The new ground classification algorithm is intended to become a more performant open source option compared to the existing algorithms in lidR/lasR.

   This new algorithm has been tested in a standalone approach, and validated with real data. The next step is to integrate it in the pipeline to have a fully open source pipeline with performances closer to the commercial software equivalents (lastools).

   To do this, two solutions are possible

   - Direct call to the pre-compiled executable from R, as done with lastools functions
   - Wrap the C++ code in R to be able to be used directly from the R pipeline

2. **Re-factorization of the pipeline into subsets of modular functions**

   The current state of the pipeline has all the sub-functions coded in a common source file. This lead to a complicated readability and maintenance of the code base. A restructuring of this code into individual source files for each function will be done to better organize the repository and have a faster maintenance of the pipeline.

3. **Integration and test of the pipeline in the MAAP platform**

   The Multi-Mission Algorithm and Analysis Platform is a data storage and computing platform put into service by ESA. This platform is available for the FRM4BIOMASS/GEO-TREES actors. The platform will be used to deploy the pipeline and eventually store the data from the project. The pipeline is, to our current knowledge, not still operative for production use. Once this is

the case, a deployment and test of the pipeline on the MAAP platform will be done. This will allow us to have a unified reference point for the storage and access to the data, and to efficiently process the outputs of the pipeline.

4. **Consolidation of the pipeline output products**

   The pipeline, in its actual form, produces several versions of DTM and CMH products. This has been done in order to compare different methods, as each version produces the product with a different algorithm. As the pipeline becomes mature enough, a unique algorithm will be selected for each product. The other products will be optionally produced, but not by default.

5. **Formal comparison and potential merge with Maryland's university pipeline**

   In the last stages of the development, a collaboration with the university of Maryland has been established in the context of the GEO-TREES project. The team from Maryland works on their own pipeline, open-sourced, but with different approaches compared to the ones used in our pipeline (notably for the DTM and CHM algorithms).

   The piepline source code and documentation can be found at: https://github.com/GEO-TREES/ALS_Panama

6. **Integration of the pipeline with the AGB estimation**

   - Production ready – Landscape statistical up-scaling (BIOMASS)

     Use the BIOMASS package to, using plot data of the same region scanned by the ALS data, calibrate the allometric statistical model use dfor the upscaling from the plot data to the landscape data provided by the CHM.

     More information on this approach and the package itself can be found on: https://umr-amap.github.io/BIOMASS/index.html

- Experimental – Individual tree based simulation (CanopyConstructor)

  Use the CanopyConstructor package, which uses the CHM and plot data to find correspondences between simulated trees and the landscape data, forming a simulated forest that can be used to generate AGB plots at any desired raste resolution.

  More information on this approach and the package itself can be found on: https://github.com/fischer-fjd/CanopyConstructor